

June 14, 2010

1. Here's some code... know what's what:

```
#FIXED X'S
X=seq(10,20,.25) Xt2=X*X Xt.5=sqrt(X) n=length(X)

#SIMULATED REGRESSION MODEL
alpha=1 beta=.5 sigma=.5 Y=alpha+beta*X+rnorm(n, 0, sigma) plot(X,Y)

#DATA STORAGE
my.data=data.frame(Y,X,Xt2,Xt.5) names(my.data)

#REGRESSION MODEL FITTING
lm.fit=lm(Y~X,data=my.data) lm.fit summary(lm.fit)

#REGRESSION FITTING OUTPUT
names(lm.fit) lm.fit$coefficients lm.fit$residuals lm.fit$fitted.values

#RESULT
plot(X,Y) abline(lm.fit$coefficients[1],lm.fit$coefficients[2])

#DIAGNOSTICS
plot(lm.fit$fitted.values,Y) plot(lm.fit$fitted.values,X)
plot(lm.fit$fitted.values,lm.fit$residuals) plot(X,lm.fit$residuals)

#PREDICTION, for  $E[Y|X]$  and  $Y|X$ , respectively
predict(lm.fit,newdata=data.frame(X=c(15,20)), interval = "confidence")
predict(lm.fit,newdata=data.frame(X=c(15,20)), interval = "prediction")
```

2. Demonstrate the uncertainty in fitting regression models (simulation).
3. Fit a simple linear model to your data (simple means a single covariate X , like above).
4. Estimate some coefficients (there's uncertainty, right?).
5. Examine the regression assumptions in your context (hmm... how will you do this?).
6. Do some prediction (there's uncertainty, right?),
7. and then give intervals for where new Y 's might be for certain levels of X .
8. Hint/Question for previous two questions: What's the difference between the two predictions above? (R seems to call one 'confidence' and one 'prediction')?
9. Last: Do X 's that are farther apart allow for better model estimation? I.e., reduced uncertainty in model fit? Show this theoretically as well as through simulation.
10. Last last: 12-5, 12-8, 12-9, 12-18